



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0104645
(43) 공개일자 2024년07월05일

(51) 국제특허분류(Int. Cl.)
G06N 3/063 (2023.01) G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 3/063 (2013.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2022-0187082
(22) 출원일자 2022년12월28일
심사청구일자 2022년12월28일

(71) 출원인
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
김세영
경상북도 포항시 남구 효자동 효자로 62 302동
203호 (효자동, 테라비아타)
김도윤
경기도 고양시 일산동구 하늘마을1로 7 하늘마을
4단지아파트
(74) 대리인
정부연

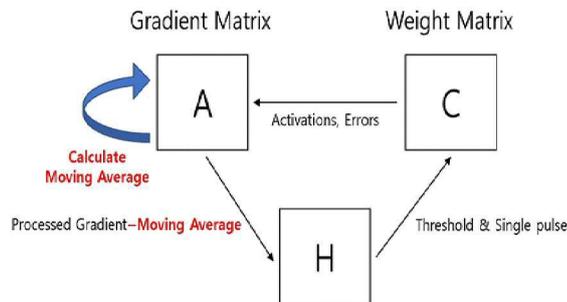
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 시냅스 어레이 소자 및 이를 이용한 인공 신경망 학습 방법

(57) 요약

본 발명의 일 실시예에 따른 시냅스 어레이 소자는 가중치 값(weight)을 나타내는 제1 시냅스 어레이와, 상기 제1 시냅스 어레이의 가중치 값의 에러 변화 값(gradient)을 전달받아 행 단위로 정제한 변화 값을 나타내는 제2 시냅스 어레이와, 상기 제2 시냅스 어레이로부터 행 단위로 정제된 변화 값을 전달받고, 이 중 임계 값(threshold)이 초과된 부분을 상기 제1 시냅스 어레이로 전달하는 제3 시냅스 어레이를 포함하며, 상기 제3 시냅스 어레이는 상기 제2 시냅스 어레이로부터 전달받은 값들의 누적 값을 평균 내어 이동 평균 값을 도출하고, 도출된 이동 평균 값을 상기 제2 시냅스 어레이로 전달하는 것을 특징으로 한다.

대표도 - 도6



명세서

청구범위

청구항 1

가중치 값(weight)을 나타내는 제1 시냅스 어레이;

상기 제1 시냅스 어레이의 가중치 값의 에러 변화 값(gradient)을 전달받아 행 단위로 정제된 변화 값을 나타내는 제2 시냅스 어레이; 및

상기 제2 시냅스 어레이로부터 행 단위로 정제된 변화 값을 전달받고, 이 중 임계 값(threshold)이 초과된 부분을 상기 제1 시냅스 어레이로 전달하는 제3 시냅스 어레이를 포함하며, 상기 제3 시냅스 어레이는 상기 제2 시냅스 어레이로부터 전달받은 값들의 누적 값을 평균 내어 이동 평균 값을 도출하고, 도출된 이동 평균 값을 상기 제2 시냅스 어레이로 전달하는 것을 특징으로 하는 시냅스 어레이 소자.

청구항 2

제1 항에 있어서,

상기 제1 시냅스 어레이 및 제2 시냅스 어레이는 아날로그(analog) 어레이 디바이스를 사용하고, 상기 제3 시냅스 어레이는 디지털(digital) 상에 위치한 것을 특징으로 하는 시냅스 어레이 소자.

청구항 3

제1 항에 있어서,

상기 제1 시냅스 어레이, 제2 시냅스 어레이 및 제3 시냅스 어레이들 사이의 학습(training) 과정이 반복될수록 상기 제2 시냅스 어레이는 '0'에 수렴하고, 변화 값은 '0'에 인접하는 것을 특징으로 하는 시냅스 어레이 소자.

청구항 4

제1 항에 있어서,

상기 제2 시냅스 어레이의 이동 평균 값 전달 시 계속적으로 이동 평균 값을 업데이트 하여 이를 오프 셋(off set)으로 설정하는 것을 특징으로 하는 시냅스 어레이 소자.

청구항 5

제1 항에 있어서,

상기 이동 평균 값은 기존의 평균 값과 새로운 제2 시냅스 어레이의 값을 특정 비율로 더해주는 형태로 계산될 수 있으며, 특정 비율은 일정하게 유지되거나 해당 값을 변동하여 수렴 정도를 조절하는 것을 특징으로 하는 시냅스 어레이 소자.

청구항 6

아날로그 어레이 디바이스인 제1 시냅스 어레이 및 제2 시냅스 어레이와 디지털 상에 위치한 제3 시냅스 어레이를 포함하는 시냅스 어레이 소자를 이용한 인공 신경망 학습 방법에 있어서,

상기 제1 시냅스 어레이의 가중치 값의 에러 변화 값(gradient)을 제2 시냅스 어레이로 전달하는 단계;

상기 제2 시냅스 어레이를 통해 상기 에러 변화 값을 행 단위로 정제하고, 해당 변화 값을 상기 제3 시냅스 어레이로 전달하는 단계; 및

상기 제3 시냅스 어레이를 통해 상기 제2 시냅스 어레이로부터 전달받은 변화 값들의 누적 값을 평균내어 이동 평균 값을 도출하고, 이동 평균 값을 다시 상기 제2 시냅스 어레이로 전달하는 단계

를 포함하는 것을 특징으로 하는 인공 신경망 학습 방법.

청구항 7

제6 항에 있어서,

상기 제2 시냅스 어레이의 초기 설정(initialization)은 대칭 점(symmetry point) 또는 그 외의 값으로 설정이 가능한 것을 특징으로 하는 인공 신경망 학습 방법.

청구항 8

제6 항에 있어서,

상기 이동 평균 값(moving average)은 아래의 <수학식 1>과 같은 방식으로 계산되는 것을 특징으로 하는 인공 신경망 학습 방법.

<수학식 1>

$$moving\ average = moving\ average * \frac{window-1}{window} + A * \frac{1}{window}$$

청구항 9

제8 항에 있어서,

상기 <수학식 1>의 윈도우(window) 값은 해당 값을 조절하여 수렴 정도를 조절할 수 있으며, 이 값은 경우에 따라 일정하거나 에포크(epoch)에 따라 이동 평균 값이나 상기 제1 시냅스 어레이 및 상기 제2 시냅스 어레이 값에 대한 함수로 지속적으로 변화 가능한 것을 특징으로 하는 인공 신경망 학습 방법.

청구항 10

제6 항에 있어서,

상기 해당 변화 값을 상기 제3 시냅스 어레이로 전달하는 단계에서

상기 이동 평균 값을 지속적으로 업데이트 하여 이를 오프셋(offset) 처리하는 것을 특징으로 하는 인공 신경망 학습 방법.

청구항 11

제10 항에 있어서,

상기 업데이트가 진행될 때마다 주기적인 감쇠를 넣어줄 수 있으며, 상기 주기적인 감쇠는 0과 1사이의 감마 파라미터(gamma parameter)를 정의하고, 이를 이동 평균 값에 곱해줌으로서 이동 평균 값을 감소시키는 방식으로 적용하는 것을 특징으로 하는 인공 신경망 학습 방법.

발명의 설명

기술 분야

[0001] 본 발명은 비대칭적 동작 경향성을 보이는 시냅스 어레이 소자에서 누적된 기울기 행렬의 평균값을 활용하여 메인 행렬 수렴 값의 오차를 감소시키는 시냅스 어레이 소자 및 이를 이용한 인공 신경망 학습 방법에 관한 기술이다.

배경 기술

[0003] 최근 들어 뉴럴 네트워크를 하드웨어적으로 구현한 뉴로모픽 소자에 대한 연구가 다양한 방향으로 진행되고 있다. 뉴로모픽 소자는 생체의 뇌신경계를 이루는 뉴런과 시냅스의 구조를 모방한 것으로, 대체로 시냅스 전에 위치한 시냅스 전 뉴런(Pre neuron), 시냅스, 시냅스 후에 위치한 시냅스 후 뉴런(Post neuron)의 구조를 갖는다. 시냅스는 뉴런과 뉴런 사이의 연결 지점으로써, 양쪽 뉴런에서 발생한 스파이크 (spike) 신호에 따라 시냅스 가중치 (synaptic weight)를 업데이트하고(updating), 이를 저장하는 (memorizing) 기능을 가진다.

[0004] 일반적으로 시냅스 소자 어레이를 사용하여 뉴럴 네트워크를 학습시킬 때에는 학습 성능을 올리기 위해 시냅스 소자에 정확한 가중치를 업데이트하는 것이 중요하다. 따라서 한 번의 업데이트 시 시냅스 소자에 업데이트되는 컨덕턴스 값이 목표 값과 일치해야한다. 그러나 ReRAM(Resistive RAM), PCM(Phase Change Memory), FeRAM(Ferroelectric RAM), ECRAM(Electrochemical RAM) 등 활발히 연구되고 있는 시냅스 소자들의 경우, 동일한 컨덕턴스 값을 갖고 있는 상태에서도 컨덕턴스의 증감 방향에 따라 한 번에 업데이트되는 컨덕턴스의 양이 달라진다. 이를 소자의 비대칭적 업데이트 특성이라고 하며, 이러한 비대칭적 업데이트 특성은 소자에 정확한 가중치 값이 저장되는 것을 방해하므로 뉴럴 네트워크 학습 성능 악화의 주요 원인이 된다. 하지만 이는 시냅스 소자의 구조 및 이에 따른 컨덕턴스 변화 기전에 따른 물리적 특성이므로, 소자의 업데이트 비대칭성을 개선하는 연구가 계속되고 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 공개특허 제10-2020-0100286호(2020.08.26)

발명의 내용

해결하려는 과제

[0007] 본 발명의 일 실시예는 누적된 값의 평균값을 계산하여 빼 주는 방식의 알고리즘의 인공 신경망 학습 방법을 시냅스 소자에 적용함으로써, 비 이상적인 소자에서 가중치가 위, 아래로 업데이트 되는 양이 동일한 지점인 대칭점을 0으로 설정해 주지 않더라도 가중치 연산의 오차가 발생하지 않도록 하는 시냅스 어레이 소자 및 이를 이용한 인공 신경망 학습 방법을 제공하고자 한다.

[0008] 본 발명의 일 실시예는 업데이트 과정을 보다 단순화시키며, 발생할 수 있는 오차를 감소시켜 아날로그 뉴럴 네트워크 가속기를 구현할 때 보다 디지털에 가까운 학습 성능을 보장하는데 도움을 줄 수 있는 시냅스 어레이 소자 및 이를 이용한 인공 신경망 학습 방법을 제공하고자 한다.

과제의 해결 수단

[0010] 본 발명의 일 실시예에 따른 시냅스 어레이 소자는 가중치 값(weight)을 나타내는 제1 시냅스 어레이와, 상기 제1 시냅스 어레이의 가중치 값의 에러 변화 값(gradient)을 전달받아 행 단위로 정제된 변화 값을 나타내는 제2 시냅스 어레이와, 상기 제2 시냅스 어레이로부터 행 단위로 정제된 변화 값을 전달받고, 이 중 임계 값

(threshold)이 초과된 부분을 상기 제1 시냅스 어레이로 전달하는 제3 시냅스 어레이를 포함하며, 상기 제3 시냅스 어레이는 상기 제2 시냅스 어레이로부터 전달받은 값들의 누적 값을 평균 내어 이동 평균 값을 도출하고, 도출된 이동 평균 값을 상기 제2 시냅스 어레이로 전달하는 것을 특징으로 한다.

- [0011] 상기 제1 시냅스 어레이 및 제2 시냅스 어레이는 아날로그(analog) 어레이 디바이스를 사용하고, 상기 제3 시냅스 어레이는 디지털(digital) 상에 위치된 것을 특징으로 하는 시냅스 어레이 소자.
- [0012] 상기 제1 시냅스 어레이, 제2 시냅스 어레이 및 제3 시냅스 어레이들 사이의 학습(training) 과정이 반복될수록 상기 제2 시냅스 어레이는 '0'에 수렴하고, 변화 값은 '0'에 인접하는 것을 특징으로 한다.
- [0013] 상기 제2 시냅스 어레이의 이동 평균 값 전달 시 계속적으로 이동 평균 값을 업데이트 하여 이를 오프 셋(offset)으로 설정할 수 있다.
- [0014] 상기 이동 평균 값은 기존의 평균 값과 새로운 제2 시냅스 어레이의 값을 특정 비율로 더해주는 형태로 계산될 수 있으며, 특정 비율은 일정하게 유지되거나 해당 값을 변동하여 수렴 정도를 조절하는 것을 특징으로 한다.
- [0015] 본 발명의 일 실시예에 따른 인공 신경망 학습 방법은 아날로그 어레이 디바이스인 제1 시냅스 어레이 및 제2 시냅스 어레이와 디지털 상에 위치된 제3 시냅스 어레이를 포함하는 시냅스 어레이 소자를 이용한 인공 신경망 학습 방법에 있어서, 상기 제1 시냅스 어레이의 가중치 값의 에러 변화 값(gradient)을 제2 시냅스 어레이로 전달하는 단계와, 상기 제2 시냅스 어레이를 통해 상기 에러 변화 값을 행 단위로 정제하고, 해당 변화 값을 상기 제3 시냅스 어레이로 전달하는 단계와, 상기 제3 시냅스 어레이를 통해 상기 제2 시냅스 어레이로부터 전달받은 변화 값들의 누적 값을 평균내어 이동 평균 값을 도출하고, 이동 평균 값을 다시 상기 제2 시냅스 어레이로 전달하는 단계를 포함하는 것을 특징으로 한다.
- [0016] 상기 제2 시냅스 어레이의 초기 설정(initialization)은 대칭 점(symmetry point) 또는 그 외의 값으로 설정이 가능한 것을 특징으로 한다.
- [0017] 상기 이동 평균 값(moving average)은 아래의 <수학식 1>과 같은 방식으로 계산되는 것을 특징으로 한다.
- [0018] <수학식 1>

$$moving\ average = moving\ average * \frac{window-1}{window} + A * \frac{1}{window}$$

- [0019]
- [0020] 상기 <수학식 1>의 윈도우(window) 값은 해당 값을 조절하여 수렴 정도를 조절할 수 있으며, 이 값은 경우에 따라 일정하거나 에포크(epoch)에 따라 이동 평균 값이나 상기 제1 시냅스 어레이 및 상기 제2 시냅스 어레이 값에 대한 함수로 지속적으로 변화 가능한 것을 특징으로 한다.
- [0021] 상기 해당 변화 값을 상기 제3 시냅스 어레이로 전달하는 단계에서 상기 이동 평균 값을 지속적으로 업데이트 하여 이를 오프셋(offset) 처리하는 것을 특징으로 한다.
- [0022] 상기 업데이트가 진행될 때마다 주기적인 감쇠를 넣어줄 수 있으며, 상기 주기적인 감쇠는 0과 1사이의 감마 파라미터(gamma parameter)를 정의하고, 이를 이동 평균 값에 곱해줌으로서 이동 평균 값을 감소시키는 방식으로 적용하는 것을 특징으로 한다.

발명의 효과

- [0025] 개시된 기술은 다음의 효과를 가질 수 있다. 다만, 특정 실시예가 다음의 효과를 전부 포함하여야 한다거나 다음의 효과만을 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [0026] 본 발명의 일 실시예에 따른 시냅스 어레이 소자 및 이를 이용한 인공 신경망 학습 방법은 누적된 값의 평균값을 계산하여 빼 주는 방식의 알고리즘의 인공 신경망 학습 방법을 시냅스 소자에 적용함으로써, 비 이상적인 소자에서 가중치가 위, 아래로 업데이트 되는 양이 동일한 지점인 대칭 점을 0으로 설정해 주지 않더라도 가중치

연산의 오차가 발생하지 않도록 하는 효과를 제공한다.

[0027] 본 발명의 일 실시예에 따른 시냅스 어레이 소자 및 이를 이용한 인공 신경망 학습 방법은 업데이트 과정을 보다 단순화시키며, 발생할 수 있는 오차를 감소시켜 아날로그 뉴럴 네트워크 가속기를 구현할 때 보다 디지털에 가까운 학습 성능을 보장하는데 도움을 줄 수 있다.

도면의 간단한 설명

- [0029] 도 1은 일반적인 인공 신경망의 구조를 간략하게 도시한 도면이다.
- 도 2는 인공 신경망을 구현한 아날로그 RPU의 예시를 나타낸 도면이다.
- 도 3은 아날로그 시냅스 소자의 업데이트 경향성을 나타낸 그래프이다.
- 도 4는 비 이상성(non-ideality)을 보완할 수 있는 알고리즘(Tiki-Taka)에 대한 구조를 설명하기 위한 도면이다.
- 도 5는 도 4의 알고리즘을 적용했을 때 A 매트릭스 및 C 매트릭스의 동작을 나타낸 시뮬레이션 결과를 도시한 것이다.
- 도 6은 본 발명의 일 실시예에 따른 알고리즘의 구조도를 나타낸 도면이다.
- 도 7 및 도 8은 도 6의 알고리즘을 순서대로 나타낸 도면이다.
- 도 9는 도 6의 알고리즘 동작의 제1 시냅스 어레이 및 제2 시냅스 어레이의 동작 경향성을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 본 발명에 관한 설명은 구조적 내지 기능적 설명을 위한 실시예에 불과하므로, 본 발명의 권리범위는 본문에 설명된 실시예에 의하여 제한되는 것으로 해석되어서는 아니 된다. 즉, 실시예는 다양한 변경이 가능하고 여러 가지 형태를 가질 수 있으므로 본 발명의 권리범위는 기술적 사상을 실현할 수 있는 균등물들을 포함하는 것으로 이해되어야 한다. 또한, 본 발명에서 제시된 목적 또는 효과는 특정 실시예가 이를 전부 포함하여야 한다거나 그러한 효과만을 포함하여야 한다는 의미는 아니므로, 본 발명의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [0031] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0032] "제1", "제2" 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위한 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.
- [0033] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결될 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다고 언급된 때에는 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 한편, 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0034] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0035] 각 단계들에 있어 식별부호(예를 들어, a, b, c 등)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 단계들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [0036] 본 발명은 컴퓨터가 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현될 수 있고, 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함

한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0037] 여기서 사용되는 모든 용어들은 다르게 정의되지 않는 한, 본 발명이 속하는 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한 이상적이거나 과도하게 형식적인 의미를 지니는 것으로 해석될 수 없다.

[0038] 이하 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 이하 도면상의 동일한 구성 요소에 대하여는 동일한 참조 부호를 사용하고, 동일한 구성 요소에 대해서 중복된 설명은 생략한다.

[0040] 아날로그 딥 뉴럴 네트워크 연산을 통한 on-device training은 뉴로모픽 연산 중 가장 많은 비중을 차지하는 행렬 연산을 메모리에서 수행할 수 있다. 이는 기존 컴퓨터 구조에서의 메모리와 연산장치가 다른 곳에 존재하여, 데이터를 옮겨와야 하는 번거로움을 해결할 수 있어, 이를 통해 저전력, 고효율의 연산이 가능하게 한다. 그러나, 이를 구현하기 위한 소자의 스펙은 매우 높아야 하며, 비이상적인 부분을 줄여야 하는데 지금까지 제시되고 있는 디바이스들은 이를 만족하지 못하므로 최근에는 알고리즘을 더 간소화시켜 저전력, 고효율 on-device training의 구현이 요구되고 있는 실정이다.

[0042] 도 1은 일반적인 인공 신경망의 구조를 간략하게 도시한 도면이다.

[0043] 도 1을 참조하면, 인공 신경망은 입력 계층(Input layer)(100), 은닉 계층(hidden layer)(110) 및 출력 계층(output layer)(120)로 구성된다. 여러 뉴런이 모인 단위를 계층(layer)라고 하며, 전결합 계층 구조는 각 계층의 모든 경우들이 연결되어 있는 구조이다. 입력 계층의 뉴런들과 출력 계층의 뉴런들이 연결될 수 있는 모든 경우의 수와 동일하게 연결이 되어 있으면 이를 전결합 계층(Fully Connected Layer)라고 한다.

[0044] 입력 계층은 입력을 받아서 다음 계층인 은닉 계층으로 넘기는 역할을 하고, 은닉 계층은 입력 계층과 연결된 전결합 계층이며, 복잡한 문제를 해결할 수 있게 하는 핵심적인 계층이라고 할 수 있다. 마지막으로 출력계층은 은닉 계층 다음에 오는 전결합 계층으로, 신경망의 외부로 출력 신호를 전달하는 데에 사용하며, 신경망의 기능은 출력 계층의 활성화 함수에 의해 결정된다.

[0045] 트레이닝 과정은 정방향 패스(forward pass) 및 역방향 패스(backward pass)로 구성되어 있으며, 들어온 인풋값은 은닉 계층(110)들을 거쳐서 출력 계층(120)으로 나오며 이를 정방향 패스라고 한다. 또한 에러(error)값의 변화 값(gradient)은 역방향 패스를 통해 각 뉴런에 전달되며, 이후 업데이트가 이루어진다.

[0047] 도 2는 인공 신경망을 구현한 아날로그 RPU(Resistive Processing Unit)의 예시를 나타낸 도면이다.

[0048] 도 2를 참조하면, 각 RPU는 하나의 가중치(weight)를 나타낼 수 있는 컨덕턴스(conductance)를 가지고 있으며, 이는 뉴럴 네트워크의 행렬 연산을 담당한다. 이는 옴의 법칙과 키르히호프의 전류 법칙에 따라서 연산이 이루어지며, 주변 디지털 회로를 통해 기타 연산이 구현된다.

[0050] 도 3은 아날로그 시냅스 소자의 업데이트 경향성을 나타낸 그래프로, 도 3a의 (i)는 이상적인 아날로그 시냅스 소자의 업데이트 경향성을 나타낸 것이고, 도 3b의 (ii)는 일반적인 아날로그 시냅스 소자의 업데이트 경향성을 나타낸 것이다.

[0051] 먼저, 도 3a의 (i)을 참조하면, 이상적인 아날로그 시냅스 소자는 +방향과 -방향에서의 업데이트 경향성이 완벽하게 대칭적이어야 하며, 매우 작은 소음(noise)와 유지(retention), 분산(variance)을 가지며, 상태의 수(number of state)가 충분히 많아야 한다.

[0052] 그러나, 현실적으로는 도 3a (ii)와 같이 + 방향과 -방향의 업데이트 경향성이 다르며, 그 외에도 3a에서 설명한 소음, 유지, 분산 및 상태의 수와 같은 부분에서 비 이상성을 가진다. 이와 같이 비선형적 특성을 가지므로,

+ 방향과 -방향의 업데이트 경향성이 대칭되는 지점인 대칭 점(symmetry point)을 찾고, 이 지점을 '0' 포인트로 설정한 후 동작을 진행시켜야 한다. 도 3b는 업데이트 경향성이 대칭되는 지점을 찾는 과정을 나타내고 있다. 임의의 지점에서 1회씩, 또는 같은 양만큼 상승 및 하강을 반복하게 될 경우, 가중치 값은 대칭 점(symmetry point)으로 수렴하게 된다. 이는 대칭 점이 아닌 곳에서는 상승 및 하강 시 업데이트 되는 양이 서로 다르기 때문에 1회의 상승 및 하강을 반복하면 특정 방향으로 이동하는데, 이는 대칭 점을 향한다.

- [0054] 도 4는 비 이상성을 보완할 수 있는 알고리즘(Tiki-Taka)에 대한 구조를 설명하기 위한 도면이다.
- [0055] 먼저, 도 4a를 참조하면, 해당 알고리즘은 메인 어레이인 C 매트릭스와, 서브 어레이이자 변화 값(gradient)을 나타내는 A 매트릭스 2개를 사용하며, 특정한 학습 비율에 따라 C 매트릭스의 변화 값을 A 매트릭스에 업데이트, 특정 에포크(epoch)마다 A 매트릭스의 일부 값을 C 매트릭스에 업데이트함으로써 이루어진다.
- [0056] 또한, 도 4b를 참조하면, 도 4a의 알고리즘에서 새로운 매트릭스를 추가하여 비 이상성에 대한 허용 오차 (tolerance)를 높인 새로운 알고리즘(Tiki-Taka v2)을 나타낸 것이다. H 매트릭스는 디지털 상에 존재하며, A 매트릭스의 값이 전달되어 업데이트되며, 해당 값이 특정 임계 값 이상일 경우 C 매트릭스로 전달되는 구조를 가지고 있다.
- [0058] 도 5는 도 4의 알고리즘을 적용했을 때 A 매트릭스 및 C 매트릭스의 동작을 나타낸 시뮬레이션 결과를 도시한 것이다.
- [0059] 도 5a를 참조하면, A 매트릭스는 비대칭한 소자에서 적절한 목표 값이 어디이던 간에 0을 향해 수렴한다. A 매트릭스는 소자의 변화 값을 나타내므로, 이는 곧 C 매트릭스가 목표 값에 수렴함을 의미한다.
- [0060] 한편, 도 5b는 A 매트릭스의 '0'점이 대칭 점과 일치하지 않게 되었을 때를 나타낸 시뮬레이션 결과이다. A 매트릭스의 트레이닝 횟수가 반복될수록 A 매트릭스는 영점이 아닌 대칭 점으로 수렴하려고 한다. 즉 변화 값이 0이 아닌 점으로 수렴한다는 의미이며, 이는 C 매트릭스의 값에 오차가 생긴다는 의미가 된다. 따라서 A 매트릭스는 대칭 점 이동 기술(symmetry point shifting technique)을 통해 영점을 대칭 점으로 맞춰주어야 한다.
- [0062] 도 6은 본 발명의 일 실시예에 따른 알고리즘의 구조도를 나타낸 도면이다.
- [0063] 도 6을 참조하면, 본 발명의 시냅스 어레이 소자는 메인 어레이인 제1 시냅스 어레이(C)와 서브 어레이인 제2 시냅스 어레이(A)로 구성된다. 또한, 제2 시냅스 어레이(A)로부터 전달받은 값들의 누적 값을 평균 낸 이동 평균 값을 도출하고, 이 이동 평균 값을 다시 제2 시냅스 어레이(A)로 전달하는 제3 시냅스 어레이(H)를 더 포함한다. 이때, 제1 시냅스 어레이(C) 및 제2 시냅스 어레이(A)는 2개의 아날로그 어레이 디바이스를 사용하며, 제3 시냅스 어레이(H)는 디지털 상에 위치시킨다.
- [0064] 제1 시냅스 어레이(C)는 가중치 값(weight)을 나타내며, 제2 시냅스 어레이(A)는 제1 시냅스 어레이(C)의 가중치 값의 예러 변화 값(gradient)을 전달받아 행별로 정제된 변화 값을 나타낸다.
- [0065] 또한, 제3 시냅스 어레이(H)는 제2 시냅스 어레이(A)로부터 행 단위로 정제된 변화 값을 전달받는다. 전달받은 변화 값 중 임계 값(threshold)이 초과된 부분을 제1 시냅스 어레이(C)로 전달하고, 매 에포크(epoch)마다 이 값을 재이동(removing)시킨다.
- [0066] 제1 시냅스 어레이(C), 제2 시냅스 어레이(A) 및 제3 시냅스 어레이(H)들 사이의 학습 과정이 반복될수록 제2 시냅스 어레이(A)는 0에 수렴하고, 변화 값은 0에 인접한다. 제1 시냅스 어레이(C)는 참 값에 수렴하게 되어 소자의 비이상성을 보완할 수 있다.
- [0067] 제3 시냅스 어레이(H)는 제2 시냅스 어레이(A)로부터 전달받은 값들의 누적 값을 평균 낸 이동 평균 값(moving average)을 도출하고, 이 이동 평균 값을 다시 제2 시냅스 어레이(A)로 다시 전달한다. 이때, 이동 평균 값은 기존의 평균 값과 새로운 제2 시냅스 어레이(A)의 값을 특정 비율로 더해주는 형태로 계산될 수 있으며, 특정 비율은 일정하게 유지되거나 해당 값을 변동하여 수렴 정도를 조절할 수 있다. 제2 시냅스 어레이의 평균 값 전달 시 계속적으로 평균 값을 업데이트 하여 이를 오프 셋(off set)으로 처리할 수 있으며, 비대칭적인 소자에서 발산하는 현상을 방지하기 위해서 업데이트 시마다 주기적인 감쇠를 넣어줄 수 있다. 상기 주기적인 감쇠는 0과

1사이의 감마 파라미터(γ parameter)를 정의하고, 이를 이동 평균 값에 곱해줌으로서 이동 평균 값을 감소시키는 방식으로 적용할 수 있다.

[0069] 이동 평균값은 소자의 특성에 따라 최적화되어야 하며, 적당한 값을 선택해 줄 경우, 매끄러운(smooth) 곡선을 그려 업데이트 값이 빠르게 한 점으로 수렴할 수 있다.

[0071] 도 7 및 도 8은 도 6의 알고리즘을 순서대로 나타낸 도면이다.

[0072] 도 7 및 도 8을 참조하면, 아날로그 어레이 디바이스인 제1 시냅스 어레이 및 제2 시냅스 어레이와 디지털 상에 위치한 제3 시냅스 어레이를 포함하는 시냅스 어레이 소자를 이용한 인공 신경망 학습 방법에 있어서, 제1 시냅스 어레이의 가중치 값의 에러 변화 값(gradient)을 제2 시냅스 어레이로 전달한다(단계 S700).

[0073] 이어서, 제2 시냅스 어레이를 통해 상기 에러 변화 값을 행 단위로 정제하고, 해당 변화 값을 상기 제3 시냅스 어레이로 전달한다(단계 S710). 그 다음 제3 시냅스 어레이를 통해 상기 제2 시냅스 어레이로부터 전달받은 변화 값들의 누적 값을 평균내어 이동 평균 값을 도출하고(단계 S720), 이동 평균 값을 다시 상기 제2 시냅스 어레이로 전달한다(단계 S730). 이때, 이동 평균 값은 아래의 <수학식 1>와 같이 계산하여 제1 시냅스 어레이(C) 또는 제3 시냅스 어레이(H)로 넘어갈 때 업데이트 해줄 수 있다.

[0074] <수학식 1>

$$\text{moving average} = \text{moving average} * \frac{\text{window}-1}{\text{window}} + A * \frac{1}{\text{window}}$$

[0075]

[0076] 상기 이동 평균 값은 기존의 평균 값과 새로운 제2 시냅스 어레이의 값을 특정한 비율로 더해준 것이며, 이 비율을 윈도우(window)라고 정의할 수 있다. 또한, 이동 평균 값은 단순한 평균 값이나 반영 비율을 여러 방식으로 변화시켜 다른 방법으로 계산하는 모든 과정을 포함할 수 있다.

[0077] <수학식 1>의 윈도우(window) 값은 해당 값을 조절하여 수렴 정도를 조절할 수 있으며, 이 값은 경우에 따라 일정하거나 에포크(epoch)에 따라 이동 평균 값이나 제1 시냅스 어레이 및 제2 시냅스 어레이 값에 대한 함수로 지속적으로 변화 가능하다.

[0078] 또한 맨 처음 제2 시냅스 어레이(A)의 초기 설정(initialization)과정은 기존 알고리즘에는 대칭 점으로 이동하여 0으로 세팅해 주어야 하지만, 도 6과 같은 알고리즘에서는 반드시 대칭 점으로 이동하여 0으로 세팅하지 않아도 된다. 즉, 본 발명의 알고리즘에서는 업데이트 양이 동일한 지점인 대칭 점을 찾아 해당 지점을 0으로 세팅해주는 과정을 생략할 수 있다.

[0080] 도 9는 도 6의 알고리즘 동작 시 제1 시냅스 어레이 및 제2 시냅스 어레이의 동작 경향성을 나타낸 것이다.

[0081] 먼저, 도 9a를 참조하면, 도 6의 알고리즘을 동작시켰을 때 제2 시냅스 어레이(A)의 동작 경향성을 나타낸다. 대칭 점으로 이동하지 않았을 때는 제2 시냅스 어레이(A)의 가중치가 0이 아닌 지점에 수렴한다. 이 때 이동 평균 값 역시 같은 지점에 수렴하므로, 이 값을 제외해주게 되면 0에 수렴함을 볼 수 있다.

[0082] 도 9b를 참조하면, 도 6의 알고리즘을 동작시켰을 때 제1 시냅스 어레이(C) 및 제2 시냅스 어레이(A)의 동작 경향성을 실제로 시뮬레이션 한 결과이다. 해당 알고리즘은 대칭 점 세팅을 하지 않아도 제1 시냅스 어레이(C)가 올바른 값으로 수렴할 수 있도록 하며, 기존 알고리즘에 비해 더 빠른 수렴이 가능하도록 한다.

[0084] 상술한 바와 같이, 본 발명은 제3 시냅스 어레이(H)에 누적된 제2 시냅스 어레이(A)의 매트릭스 값들을 이동 평균 값(moving average)라고 부르는 평균 값을 계산하여 이를 오프셋으로 설정함으로써 대칭 점(symmetry point)을 맞추는 과정을 생략할 수 있도록 한다. 이를 통해 알고리즘을 더욱 간단하게 만들어 인공 신경망의 학습과정을 더욱 용이하게 할 수 있다.

부호의 설명

[0086]

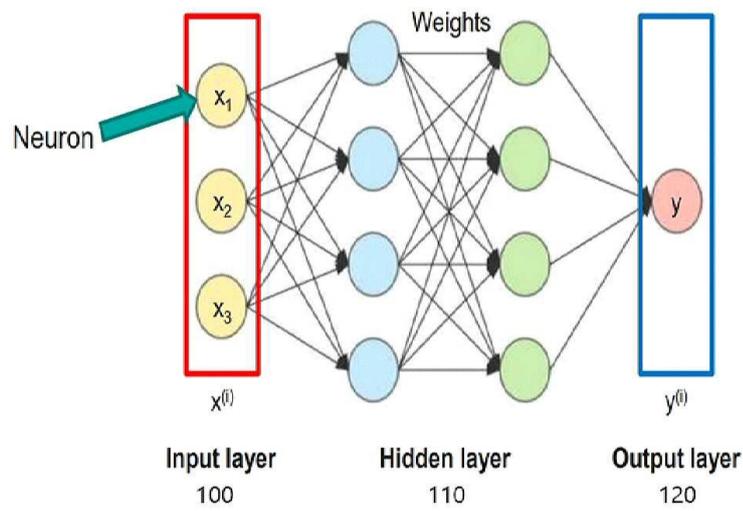
100 : 입력 계층 110 : 은닉 계층

120 : 출력 계층 A : 제2 시냅스 어레이

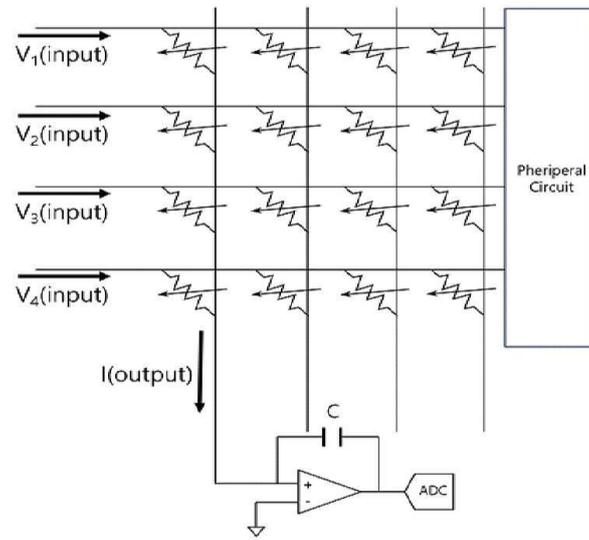
C : 제1 시냅스 어레이 H : 제3 시냅스 어레이

도면

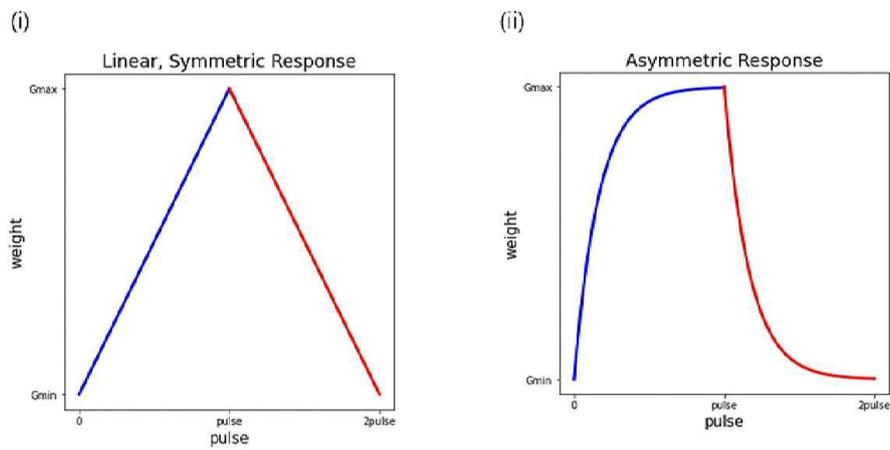
도면1



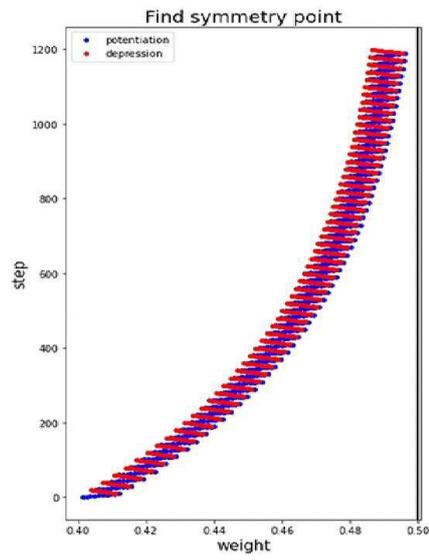
도면2



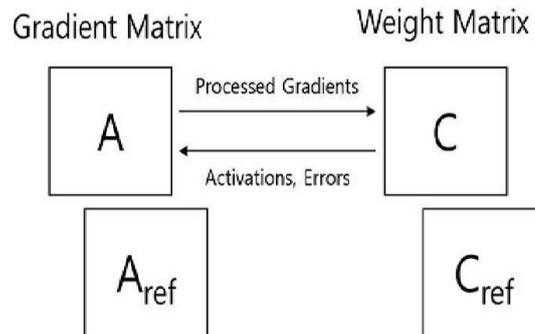
도면3a



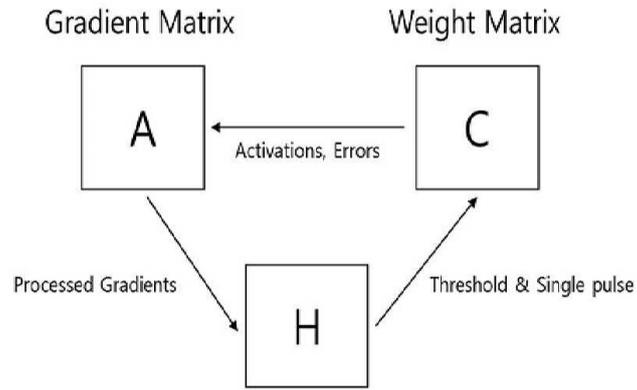
도면3b



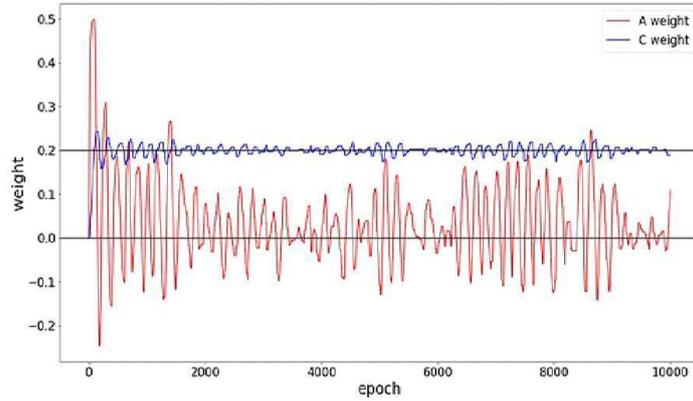
도면4a



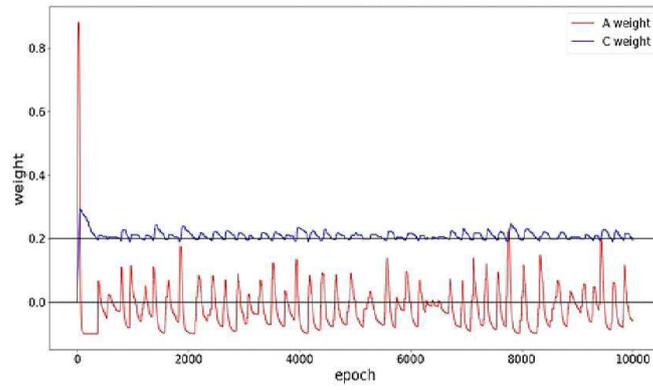
도면4b



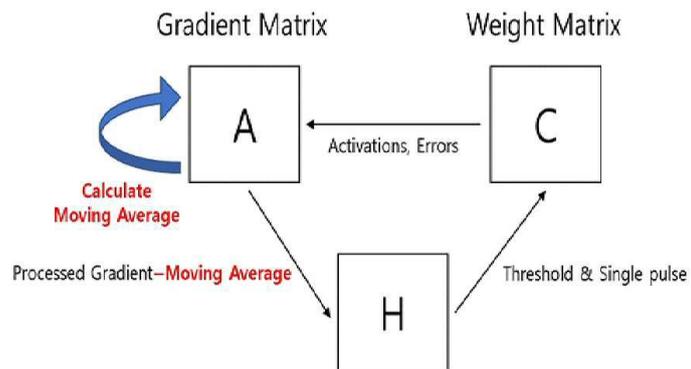
도면5a



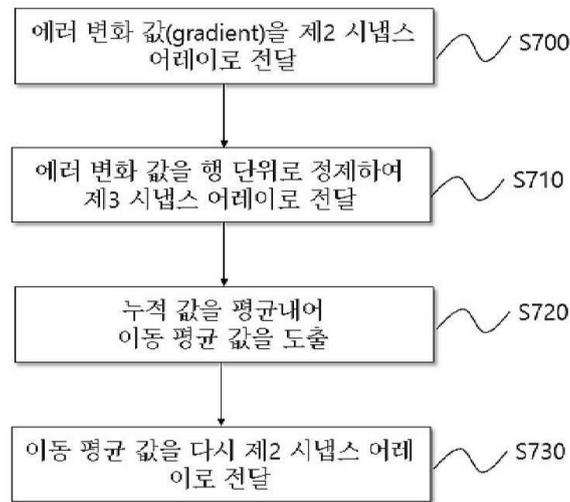
도면5b



도면6



도면7

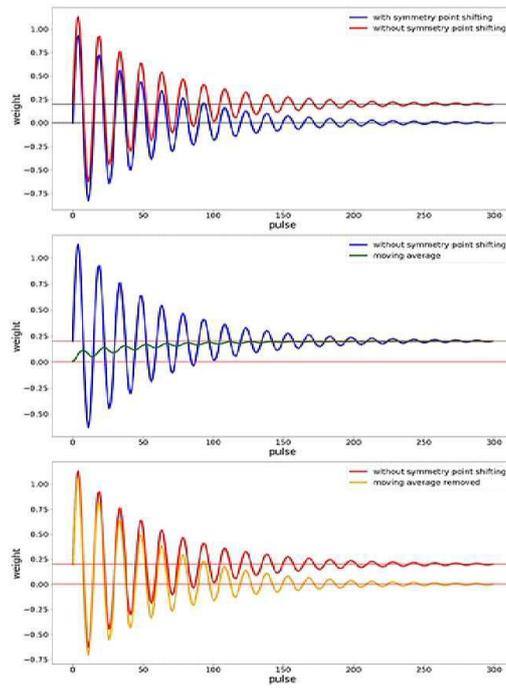


도면8

```

Initialize A to random values
Initialize C to random values
k = 0
t = 0
ns = Number of rows
ns = Number of Cycle Update
window = Large number
For each data:
  y = (yA + C)x
  y = (yA + C)Tδ
  A update
  moving average = moving average *  $\frac{window-1}{window}$  + A *  $\frac{1}{window}$ 
  k = mod(k + 1, ns)
  If (k=0)
    u(t) = [0, 0, 0 ..., 1, ..., 0] (t th value is 1)
    v = Au
    C update
    t = mod(t + 1, ns)
  end
end
  
```

도면9a



도면9b

